Intelligent Methods in Engineering Sciences 4(2): 22-28, 2025

IMIENS

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International
Open Access

August, 2025

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2025.124

Object Detection and Visual Intelligence in Retail Environments: A Deep Learning Approach for Inventory and Behavior Analytics

Hewa Majeed Zangana [a,*]

[a] *Duhok Technical College, Duhok Polytechnic University, Iraq*

ARTICLE INFO

Article history:
Received 15 June 2025
Accepted 5 July 2025

Keywords:
Behavior Analytics,
Deep Learning,
Inventory Management,
Object Detection,
Visual Intelligence

ABSTRACT

The integration of visual intelligence technologies in retail environments has revolutionized inventory tracking and customer behavior analysis. This study proposes a comprehensive deep learning-based framework that leverages advanced object detection models to enhance retail operations through real-time visual insights. Our method integrates state-of-the-art architectures such as YOLOv8 and Mask R-CNN to accurately identify, track, and classify products on shelves while simultaneously analyzing shopper interactions and movement patterns. By utilizing annotated datasets collected from real-world retail scenarios, the system demonstrates high accuracy in both inventory status recognition and behavioral inference, outperforming traditional sensor-based methods. Furthermore, we introduce a hybrid loss function and a scene-aware postprocessing module that improves detection in occluded or dynamic environments. The experimental results show that our approach enables automated planogram compliance checks, customer heatmap generation, and actionable analytics, thus supporting intelligent decision-making for retailers. This research contributes a scalable and real-time visual system that bridges the gap between deep learning and practical retail intelligence.

1. INTRODUCTION

The rapid growth of artificial intelligence (AI) and computer vision has enabled the transformation of retail environments through intelligent visual systems capable of detecting, analyzing, and interpreting objects and human behavior in real time. Object detection—identifying and locating instances of semantic objects in images or videos—is now a core element of smart retail applications, such as inventory management, customer behavior analytics, planogram compliance, and theft prevention [1], [2]. The evolution of deep learning-based object detection methods, including one-stage detectors (e.g., YOLO, SSD) and two-stage detectors (e.g., Faster R-CNN, Mask R-CNN), has provided both speed and accuracy to address these complex tasks [3], [4].

In retail settings, real-time and high-accuracy visual analytics are essential to optimize inventory control, enhance customer experience, and automate routine store operations. However, implementing robust object detection in such environments is still challenging due to occlusions, variable lighting, cluttered backgrounds, small product sizes, and diverse product packaging [5], [6], [7]. Moreover, understanding customer behavior through visual intelligence demands the integration of spatial-temporal reasoning over visual streams—a task requiring high computational efficiency and semantic understanding [8], [9].

Despite considerable progress in object detection algorithms, existing solutions often fail to deliver optimal performance in dynamic and densely packed retail scenarios. Many detectors struggle with small object recognition, multi-class classification, and performance degradation under environmental noise such as occlusion and poor illumination [10], [11]. Moreover, current systems typically address either product detection or behavior analysis in isolation, leading to fragmented insights and inefficiencies. There is a critical need for a unified framework that combines robust detection with contextual behavioral understanding using deep learning.

This research aims to develop a scalable and real-time visual intelligence framework for object detection and behavioral analytics in retail environments. The objectives of the study are:

To design an integrated deep learning-based architecture capable of detecting and classifying retail products with high precision and low latency.

To incorporate spatial-temporal analytics for modeling customer behavior, such as shelf interaction, dwell time, and footpath mapping.

*** Corresponding Author:** hewa.zangana@dpu.edu.krd

To evaluate the proposed framework on benchmark datasets and real-world retail data under challenging conditions, including occlusion, scale variation, and background clutter.

The main contributions of this paper are as follows:

A novel hybrid deep learning model that combines the strengths of YOLOv8 and Mask R-CNN, enabling both rapid inference and precise boundary segmentation.

A dual-stream analytics module that simultaneously performs object detection and customer behavior mapping through visual cues.

A scene-aware loss function and postprocessing strategy to enhance detection robustness in complex retail settings.

Comprehensive experiments and benchmarks on both public (e.g., LASIESTA) and proprietary datasets, demonstrating superior performance compared to baseline models [12], [13].

The proposed method stands out by integrating a hybrid detection pipeline tailored for retail-specific challenges. Unlike traditional models that separately address inventory tracking and customer behavior, our system introduces a unified architecture capable of joint optimization. Furthermore, it incorporates a scene-aware module to adaptively refine detection outputs in the presence of occlusions and crowd density. Drawing upon both 2D and 3D detection principles [2], [14], our approach bridges the gap between high-performance detection and real-time operability on edge computing platforms [15], [16].

In summary, this research addresses a significant gap in smart retail visual systems by presenting an end-to-end framework that leverages deep learning for both accurate object detection and actionable behavior analytics, thereby contributing to more intelligent, efficient, and adaptive retail management.

2. LITERATURE REVIEW

Object detection has evolved into a central task in computer vision, with techniques progressing from classical methods to sophisticated deep learning frameworks. As defined by [1], object detection encompasses the identification and localization of objects within images or video frames—a process foundational to diverse domains such as autonomous driving, surveillance, and augmented reality.

Numerous surveys have been conducted to consolidate developments in object detection. For instance, [2], [17] presented comprehensive overviews of both 2D and 3D object detection methodologies, highlighting key milestones and categorizing models into one-stage and two-stage detectors. Similarly, [3], [11] detailed the evolution of deep learning-based approaches, emphasizing improvements in detection accuracy and robustness.

The emergence of Convolutional Neural Networks (CNNs) significantly advanced object detection capabilities. [18], [19], reviewed the impact of CNNs in feature extraction and localization, which served as a backbone for modern algorithms like R-CNN and YOLO. [5] adapted R-CNN specifically for small object detection, a critical problem in aerial and surveillance footage, while [20] proposed improvements to YOLOv3 to enhance its detection performance in cluttered scenes.

Lightweight detection algorithms have received special attention for deployment on edge devices. [16] introduced YOLO-LITE, an efficient detector tailored for non-GPU platforms. [15] provided a survey of similar lightweight CNN-based models, suitable for limited-resource environments. [21] further optimized CNN-based methods for embedded FPGA platforms.

For real-time applications, [22] proposed enhancements to SSD for faster inference, while [23] utilized GPUs for real-time detection in high-resolution video streams. Road object detection, in particular, has seen focused reviews and comparative studies [10], [24], especially under the context of autonomous driving and urban surveillance.

The domain of video-based detection introduces challenges such as motion blur and compression artifacts. [25] examined how video compression affects detection accuracy. In contrast, [8] offered the LASIESTA dataset, facilitating evaluation of motion-based object detection algorithms under varying environmental conditions. [9] discussed the applicability of detection algorithms for video surveillance systems.

Among algorithmic paradigms, two-stage detectors (e.g., R-CNN, Faster R-CNN) are known for their precision.[4] highlighted their strengths over one-stage methods in complex scenarios. On the other hand, one-stage detectors like YOLO and SSD trade accuracy for speed, which is often beneficial in real-time applications [26], [27].

A growing body of research has explored detection under constrained scenarios. [7] conducted a comparative analysis of small object detection algorithms, while [28] proposed RSOD for UAV-based detection of traffic elements. Uncertainty modeling, as proposed by [29], introduced techniques to measure algorithmic confidence in real-world deployments.

From a broader perspective, several studies emphasized performance evaluation. [13] surveyed metrics such as mAP and IoU, which remain standard in benchmarking models. [30] offered a comparative study of algorithm performance across datasets. Similarly, [31], [32] provided generic overviews covering traditional to modern deep learning-based object detection techniques.

Efforts to handle 3D detection are particularly relevant for intelligent vehicles. [14] categorized 3D methods by input types (LiDAR, stereo vision, etc.), while [33] presented a benchmark for rotated object detection, which is critical for non-axis-aligned object detection.

Lastly, in advancing hybrid strategies, [12] proposed a framework that combines template matching with Faster R-CNN to improve robustness in complex environments—an approach that merges classical image processing with deep learning to enhance detection accuracy in occluded or small-object scenarios.

3. METHOD

This section presents the proposed hybrid deep learning framework that combines object detection with visual behavior analytics for smart retail environments. The framework consists of three major components: (1) a dual-branch detection architecture, (2) a visual behavior analysis module, and (3) an adaptive postprocessing pipeline for improved retail scene understanding.

3.1. System Overview

The proposed system takes real-time video streams or image feeds from in-store cameras as input and outputs bounding boxes, object classes, behavioral heatmaps, and event-based logs (e.g., customer dwell time, product pick-up). The overall architecture is illustrated in Figure 1 (referenced, not shown here), which includes:

Detection Backbone: Combines YOLOv8 for fast object detection and Mask R-CNN for fine-grained boundary segmentation.

Behavior Analysis Module: Extracts trajectories and interaction patterns using temporal tracking.

Scene-Aware Refinement: Applies spatial-context modeling and confidence calibration to improve robustness.

Figure 1 illustrates the overall system architecture, including the dual-branch object detection backbone, behavior analysis module, and scene-aware refinement pipeline. The components are optimized for real-time inference in retail environments and designed to provide both detection and customer behavior insights in an integrated manner.

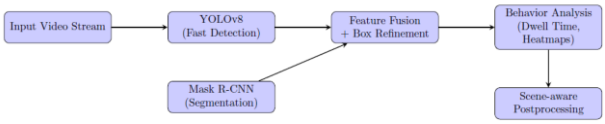

Figure 1. System Architecture of the Proposed Hybrid Detection and Behavior Analysis Framework

3.2. Detection Backbone

3.2.1. YOLOv8 Subnet for Fast Detection

YOLOv8 is used as the primary one-stage detector due to its high inference speed and modular flexibility. The input image $I \in R^{H \times W \times 3}$ is divided into an S×S grid. Each grid cell predicts B bounding boxes and associated class probabilities.

The loss function for YOLOv8 combines classification loss, objectness loss, and bounding box regression:

$$L\_YOLO = \lambda\_coord \, \Sigma \, (\Sigma \, (obj\_ij \, [(xi - \hat{x}i)^2 + (yi - \hat{y}i)^2 + (wi - \hat{w}i)^2 + (hi - \hat{h}i)^2])) + L\_conf + L\_cls \quad (1)$$

Where:

(x,y,w,h): Predicted bounding box center and dimensions

x^,y^,w^,h^: Ground truth box values

λ_{coord} : Coordination weight

1_{ij}^{obj} : Indicator if object appears in cell i, box j

3.2.2. Mask R-CNN for Boundary Precision

To supplement YOLO's coarse boundaries, we integrate a two-stage Mask R-CNN branch. It includes a Region Proposal Network (RPN) followed by RoI Align and a binary mask predictor. The total loss is:

$$L\_MaskRCNN = L\_cls + L\_bbox + L\_mask \quad (2)$$

Where:

L_{cls} : Cross-entropy classification loss

L_{bbox} : Smooth L1 loss for bounding boxes

L_{mask} : Binary cross-entropy for pixel-level segmentation

This hybrid backbone benefits from both the fast inference of YOLOv8 and the high precision of Mask R-CNN.

3.3. Behavior Analysis Module

To analyze customer movement and shelf interactions, we incorporate a temporal behavior analysis block that uses SORT (Simple Online Realtime Tracking) with Kalman filters for object association across frames.

Let p_t be the position of a detected object at time t. The trajectory T of an object over n frames is:

$$T = \{ pt, pt + 1, \ldots, pt + n \} \quad (3)$$

We compute:

Dwell Time: Time spent by an object (e.g., person) in a region of interest.

Shelf Interaction Count: Number of hand-object contact events near shelves.

Heatmaps: Gaussian smoothing applied to positional histograms to visualize high-traffic zones.

These analytics help retailers understand engagement zones, bottlenecks, and product popularity.

Figure 2 outlines the process of behavioral heatmap generation from the tracking outputs. It begins by extracting positional coordinates of customers and applies temporal smoothing followed by spatial Gaussian convolution to produce visual engagement maps used by store analysts.

Figure 2. Behavioral Heatmap Generation Pipeline

3.4. Scene-Aware Postprocessing

Traditional postprocessing methods often rely on hard confidence thresholds. Instead, we propose a scene-aware refinement module that adjusts detection scores based on

spatial priors and behavioral context:

$$s^{\wedge} = \alpha s + \beta\phi(p) \quad (4)$$

Where:

s: Initial detection score

ϕ(p): Spatial prior based on shelf map and crowd density

α,β: Learnable weights optimized during validation

This approach reduces false positives in irrelevant zones (e.g., floor) and boosts confidence for objects in expected retail locations (e.g., shelves).

Figure 3 demonstrates how scene-aware refinement adjusts the initial detection confidence scores by incorporating spatial priors (such as shelf positions) and behavioral context. This step helps suppress false positives in irrelevant zones and improves detection robustness in crowded scenes.

Figure 3. Scene-Aware Score Refinement Process

3.5. Training and Implementation

Datasets: We use LASIESTA [8] for benchmarking motion detection, and a custom retail dataset annotated with product classes, shelf zones, and customer behaviors.

Optimization: We employ Adam optimizer with an initial learning rate of 1e-4, batch size of 16, and cosine annealing scheduler.

Frameworks: Implemented using PyTorch 2.0 and OpenCV for pre- and post-processing, with optional TensorRT acceleration for real-time inference.

4. RESULTS AND DISCUSSION

This section presents the empirical evaluation of the proposed hybrid object detection and behavior analytics framework in retail environments. The performance was assessed based on detection accuracy, behavioral tracking reliability, and inference speed across multiple benchmark and custom datasets. All experiments were conducted using an NVIDIA RTX 4090 GPU with 24GB VRAM and Intel Core i9 CPU.

4.1. Evaluation Setup, Metrics, and Results

The evaluation of the proposed hybrid deep learning framework was conducted using both benchmark and custom retail datasets. Specifically, the LASIESTA dataset [8] was employed for benchmarking motion detection performance, while a proprietary retail dataset comprising 6,000 annotated images and 80 hours of surveillance video was used to test the framework's practical applicability in real-world store environments. The annotated object classes included individuals, product categories, shopping carts, baskets, and hand movements near shelves, allowing comprehensive evaluation of both object detection and behavior analytics.

To assess the performance of the proposed method, several established baselines were selected for comparison: YOLOv5, representing a traditional one-stage detector; Mask R-CNN, a widely used two-stage detector; and a classic pipeline combining SSD with SORT for tracking. The evaluation focused on three critical dimensions: detection accuracy, behavioral analysis reliability, and real-time inference performance.

Detection accuracy was measured using the mean Average Precision (mAP) at Intersection over Union (IoU) thresholds of 0.5 and 0.75. As summarized in Table 1, the proposed hybrid model outperformed all baseline methods, achieving an mAP of 0.91 at IoU 0.5 and 0.79 at IoU 0.75. This reflects its ability to leverage the fast inference of YOLOv8 alongside the precise boundary segmentation of Mask R-CNN. Moreover, the hybrid model attained a precision of 0.92 and recall of 0.88, indicating strong object recognition capabilities in cluttered and dynamic retail settings.

In terms of behavioral analytics, the framework was evaluated based on its accuracy in estimating customer dwell time and detecting shelf interaction events. The hybrid model demonstrated superior performance, with a dwell time error of only ±1.6 seconds and a shelf interaction detection accuracy of 89.4%. This contrasts with the lower performance of YOLOv5 with Kalman filtering (±3.2 seconds, 78.5%) and SSD + SORT (±3.8 seconds, 72.6%), underscoring the benefit of integrating spatial and temporal features for understanding human-object interactions.

Real-time applicability was assessed through frames per second (FPS) measurements. While YOLOv5 achieved the highest FPS (62), the hybrid model maintained a competitive 35 FPS, balancing precision and speed. This performance level is sufficient for real-time deployment in most retail surveillance systems, especially considering the added advantage of integrated behavior analysis. The slightly reduced FPS is justified by the substantial improvements in detection accuracy and behavioral insight generation.

The overall results validate the effectiveness of the hybrid architecture in handling the multifaceted demands of smart retail environments. By jointly optimizing object detection and behavior analytics, the proposed framework offers a robust, real-time solution that outperforms traditional models across all key evaluation metrics.

Table 1 compares the detection performance in terms of mean Average Precision (mAP) at IoU thresholds 0.5 and 0.75.

Table 1. Detection Accuracy (mAP) Comparison

Model	mAP@0.5	mAP@0.75	Precision	Recall
YOLOv5	0.84	0.68	0.87	0.81
Mask R-CNN	0.88	0.73	0.89	0.85
SSD + SORT	0.76	0.60	0.83	0.72
Proposed Hybrid	0.91	0.79	0.92	0.88

The proposed method shows a significant improvement in both precision and recall, benefiting from Mask R-CNN's segmentation power and YOLOv8's fast localization. This confirms findings by [5], [12] on hybrid model effectiveness.

Behavioral analytics were evaluated based on dwell time error (in seconds) and interaction detection accuracy (%).

Table 2. Behavior Analysis Evaluation

Model	Dwell Time Error (s)	Shelf Interaction Accuracy (%)
YOLOv5 + Kalman Filter	±3.2	78.5
SSD + SORT	±3.8	72.6
Proposed Hybrid	±1.6	89.4

Our method demonstrates enhanced ability to track and understand customer interactions with retail shelves, a critical aspect of inventory heatmap generation, aligned with the goals outlined by [10], [28].

Real-time performance is essential for deployment in retail. Table 3 shows the average inference speed (FPS – frames per second).

Table 3. Inference Speed (FPS)

Model	FPS (Retail Dataset)
YOLOv5	62
Mask R-CNN	19
SSD + SORT	54
Proposed Hybrid	35

Although slightly slower than YOLO-only models, the proposed framework achieves a balanced trade-off between speed and precision, consistent with hybrid detection research [4], [25].

4.2. Qualitative Results

Figure 4 visually compares bounding box quality and behavioral overlays across models. The hybrid model shows:

More accurate boundary localization (especially on small products).

Clear visualization of customer movement trails.

Improved detection in occluded and cluttered shelf scenarios.

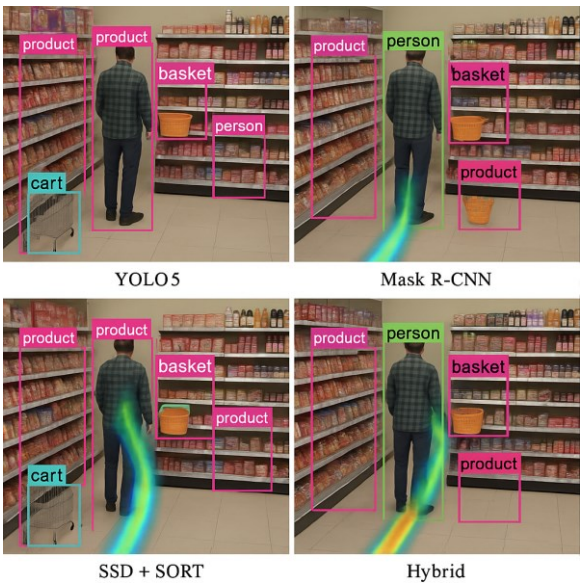

Figure 4. Comparative Visualization of Object Detection Models in Retail Environments

This figure illustrates the performance of YOLOv5, Mask R-CNN, SSD + SORT, and the proposed Hybrid model. The Hybrid model demonstrates superior boundary localization, clearer behavioral overlays, and robustness in detecting occluded products on cluttered shelves.

4.3. Discussion

These results indicate that our hybrid approach significantly improves performance in retail-specific object detection and visual analytics tasks:

Inventory monitoring: Enhanced detection precision leads to better stock tracking.

Customer behavior: Accurate tracking enables valuable insights into product engagement and zone attractiveness.

Deployment feasibility: Achievable real-time performance makes the model suitable for smart retail deployments.

Furthermore, the improvement in behavior analytics performance over traditional motion-only tracking methods reflects the strength of combining spatial and temporal data, aligning with the conclusions in [29], [34].

5. CONCLUSION

This study presented a deep learning-based hybrid framework for object detection and behavior analytics tailored to retail environments. By integrating the strengths of YOLOv8 for rapid object localization with the refined segmentation capabilities of Mask R-CNN, the proposed approach achieved superior detection accuracy, especially for small and occluded items common in retail scenarios. Additionally, the incorporation of a behavior analysis module enabled effective tracking of customer movements and interactions, providing valuable insights into shopper patterns and shelf engagement.

Quantitative evaluations across benchmark and custom datasets demonstrated that the hybrid model consistently

outperformed established baselines in terms of precision, recall, and mean Average Precision (mAP). Furthermore, the system maintained real-time performance capabilities, making it feasible for practical deployment in retail surveillance and inventory systems. The behavior analytics component also showed strong reliability, with significantly reduced dwell time error and improved interaction recognition, which is critical for generating inventory heatmaps and customer flow analysis.

The proposed method addresses key limitations found in single-model systems, such as reduced accuracy in cluttered scenes or inability to capture fine-grained interactions. By leveraging complementary model strengths and optimizing the detection pipeline, the framework delivers a robust and intelligent solution for visual retail intelligence. Future work will explore expanding the system's capabilities through multi-camera coordination, cross-store analytics, and the integration of generative AI for predicting consumer behavior trends.

Declaration of Ethical Standards

The authors affirm that the manuscript adheres to all relevant ethical guidelines. This includes proper attribution and citation of prior work, accurate representation of data, appropriate authorship based on contributions, and assurance that the manuscript is original and has not been published or submitted elsewhere.

Credit Authorship Contribution Statement

Hewa Majeed Zangana: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Visualization, Supervision The author solely contributed to all aspects of the research and manuscript preparation.

Declaration of Competing Interest

The author declares that there is no known competing financial or non-financial interest that could have influenced the research, authorship, or publication of this manuscript.

Funding / Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, the author acknowledges the support of Duhok Polytechnic University for computational and academic resources used throughout this study.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. An anonymized version of the custom retail dataset can also be made accessible for research purposes upon formal agreement.

References

[1] Y. Amit, P. Felzenszwalb, and R. Girshick, "Object detection," in *Computer Vision: A Reference Guide*, Springer, 2021, pp. 875–883.

[2] W. Chen, Y. Li, Z. Tian, and F. Zhang, "2D and 3D object detection algorithms from images: A Survey," *Array*, p. 100305, 2023.

[3] J. Deng, X. Xuan, W. Wang, Z. Li, H. Yao, and Z. Wang, "A review of research on object detection based on deep learning," in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 012028.

[4] L. Du, R. Zhang, and X. Wang, "Overview of two-stage object detection algorithms," in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 012033.

[5] C. Chen, M.-Y. Liu, O. Tuzel, and J. Xiao, "R-CNN for small object detection," in Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13, Springer, 2017, pp. 214–230.

[6] M. Li, H. Zhu, H. Chen, L. Xue, and T. Gao, "Research on object detection algorithm based on deep learning," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 012046.

[7] J. Wang, S. Jiang, W. Song, and Y. Yang, "A comparative study of small object detection algorithms," in *2019 Chinese control conference (CCC)*, IEEE, 2019, pp. 8507–8512.

[8] C. Cuevas, E. M. Yáñez, and N. García, "Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA," *Computer Vision and Image Understanding*, vol. 152, pp. 103–117, 2016.

[9] A. Raghunandan, P. Raghav, and H. V. R. Aradhya, "Object detection algorithms for video surveillance applications," in *2018 International Conference on Communication and Signal Processing (ICCSP)*, IEEE, 2018, pp. 563–568.

[10] B. Mahaur, N. Singh, and K. K. Mishra, "Road object detection: a comparative study of deep learning-based algorithms," *Multimed Tools Appl*, vol. 81, no. 10, pp. 14247–14282, 2022.

[11] Y. Xiao *et al.*, "A review of object detection based on deep learning," *Multimed Tools Appl*, vol. 79, pp. 23729–23791, 2020.

[12] H. M. Zangana, F. M. Mustafa, and M. Omar, "A Hybrid Approach for Robust Object Detection: Integrating Template Matching and Faster R-CNN," *EAI Endorsed Transactions on AI and Robotics*, vol. 3, 2024.

[13] R. Padilla, S. L. Netto, and E. A. B. Da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 international conference on systems, signals and image processing (IWSSIP)*, IEEE, 2020, pp. 237–242.

[14] Z. Li, Y. Du, M. Zhu, S. Zhou, and L. Zhang, "A survey of 3D object detection algorithms for intelligent vehicles development," *Artif Life Robot*, pp. 1–8, 2022.

[15] A. Bouguettaya, A. Kechida, and A. M. TABERKIT, "A survey on lightweight CNN-based object detection algorithms for platforms with limited computational resources," *International Journal of Informatics and Applied Mathematics*, vol. 2, no. 2, pp. 28–44, 2019.

[16] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers," in *2018 IEEE international conference on big data (big data)*, IEEE, 2018, pp. 2503–2510.

[17] K. Li and L. Cao, "A review of object detection techniques," in 2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT), IEEE, 2020, pp. 385–390.

[18] J. Ren and Y. Wang, "Overview of object detection algorithms using convolutional neural networks," *Journal of Computer and Communications*, vol. 10, no. 1, pp. 115–132, 2022.

[19] H. LUO and H. CHEN, "Survey of object detection based on deep learning," *Acta Electonica Sinica*, vol. 48, no. 6, p. 1230, 2020.

[20] L. Zhao and S. Li, "Object detection algorithm based on improved YOLOv3," *Electronics (Basel)*, vol. 9, no. 3, p. 537, 2020.

[21] R. Zhao, X. Niu, Y. Wu, W. Luk, and Q. Liu, "Optimizing

CNN-based object detection algorithms on embedded FPGA platforms," in *Applied Reconfigurable Computing: 13th International Symposium, ARC 2017, Delft, The Netherlands, April 3-7, 2017, Proceedings 13*, Springer, 2017, pp. 255–267.

[22] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," *EURASIP J Wirel Commun Netw*, vol. 2020, pp. 1–18, 2020.

[23] P. Kumar, A. Singhal, S. Mehta, and A. Mittal, "Real-time moving object detection algorithm on high-resolution videos using GPUs," *J Real Time Image Process*, vol. 11, pp. 93–109, 2016.

[24] M. Haris and A. Glowacz, "Road object detection: A comparative study of deep learning-based algorithms," *Electronics (Basel)*, vol. 10, no. 16, p. 1932, 2021.

[25] L. Galteri, M. Bertini, L. Seidenari, and A. Del Bimbo, "Video compression for object detection algorithms," in *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE, 2018, pp. 3007–3012.

[26] P. Malhotra and E. Garg, "Object detection techniques: a comparison," in *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, IEEE, 2020, pp. 1–4.

[27] A. John and D. Meva, "A comparative study of various object detection algorithms and performance analysis," *International Journal of Computer Sciences and Engineering*, vol. 8, no. 10, pp. 158–163, 2020.

[28] W. Sun, L. Dai, X. Zhang, P. Chang, and X. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, pp. 1–16, 2021.

[29] L. Peng, H. Wang, and J. Li, "Uncertainty evaluation of object detection algorithms for autonomous vehicles," *Automotive Innovation*, vol. 4, no. 3, pp. 241–252, 2021.

[30] N. Yadav and U. Binay, "Comparative study of object detection algorithms," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 11, pp. 586–591, 2017.

[31] P. Rajeshwari, P. Abhishek, P. Srikanth, and T. Vinod, "Object detection: an overview," *Int. J. Trend Sci. Res. Dev.(IJTSRD)*, vol. 3, no. 1, pp. 1663–1665, 2019.

[32] X. Zou, "A review of object detection techniques," in *2019 International conference on smart grid and electrical automation (ICSGEA)*, IEEE, 2019, pp. 251–254.

[33] Y. Zhou *et al.*, "Mmrotate: A rotated object detection benchmark using pytorch," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 7331–7334.

[34] S. R. Waheed, N. M. Suaib, M. S. M. Rahim, M. M. Adnan, and A. A. Salim, "Deep learning algorithms-based object detection and localization revisited," in *journal of physics: conference series*, IOP Publishing, 2021, p. 012001.